Article:
Husband, E. Matthew and Bovolenta, Giulia orcid.org/0000-0003-4139-6446 (2020) Prediction failure blocks the use of local semantic context. Language Cognition and Neuroscience. ISSN 2327-3801

https://doi.org/10.1080/23273798.2019.1651881

Prediction failure blocks the use of local semantic context

E. Matthew Husband[1] and Giulia Bovolenta[2]

[1]University of Oxford

[2]University of Cambridge

Mailing address:

E. Matthew Husband

St. Hugh's College

St. Margaret's Rd.

Oxford OX2 6LE

United Kingdom

Email: matthew.husband@ling-phil.ox.ac.uk (Husband), gb507@cam.ac.uk (Bovolenta)

Running head: PREDICTION FAILURE AND SEMANTIC CONTEXT

Abstract

Before accumulation of recent experimental evidence, prediction was thought to be too prone to failure and thus too costly for language comprehension. Although prediction is now widely assumed, questions about the costs of prediction failure and recovery still remain. An event-related potentials study using highly constraining Italian sentence contexts addressed these questions. It manipulated how predictive local contexts were for target nouns after cueing comprehenders to the status of global sentential predictions with article gender congruence. Predictive local contexts reduced target noun N400 amplitude when the preceding article's gender was congruent with global predictions, but not when gender was incongruent. This suggests that prediction failure impeded the facilitative use of local context for target nouns. Predictive local contexts following gender incongruence also elicited a broader late frontal positivity on target nouns, suggesting further recovery difficulties. Prediction failures, therefore, are not cost-free, and recovery from these failures requires further consideration.

Keywords: gender agreement; N400; late frontal positivity; prediction; sentence context

Introduction

A well-established finding in sentence processing is that semantic context can be used to facilitate lexical access (Tulving & Gold, 1963), and investigations over the last decade have argued that a substantial source of this facilitation comes from predictive mechanisms (Dikker, Rabagliati, & Pylkkänen, 2009; Lau, Holcomb, & Kuperberg, 2013; Federmeier, 2007; Staub, 2015). Prediction is thought to enhance recognition (Fischler & Bloom, 1979; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985), reduce the probability and duration of fixations in reading (Ashby, Rayner, & Clifton, 2005; Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996; Zola, 1984), and reduce the amplitude of the N400 (Federmeier & Kutas, 1999; Kutas & Hillyard, 1980, 1984; Wlotko & Federmeier, 2013). While many studies of these facilitatory effects were consistent with both a predictive or rapid integration account of language comprehension, recent innovations in experimental design have provided unambiguous evidence in support of a prediction mechanism (DeLong, Urbach, & Kutas, 2005; Foucart, Ruiz-Tada, & Costa, 2015; Otten, Nieuwland, & van Berkum, 2007; Szewczyk & Schriefers, 2013; van Berkum, et al., 2005; Wicha, Moreno, & Kutas, 2004). In part because of such evidence, prediction has been rapidly adopted as a core mechanism in language comprehension (Christiansen & Chater, 2016; Dell & Chang, 2014; Pickering & Garrod, 2007, 2013).

The idea that prediction could play an important role in language comprehension has not always been so clear. Before the recent accumulation of unambiguous experimental support, there was significant resistance to

predictive mechanisms in language comprehension from both conceptual and experimental perspectives (Fischler & Bloom, 1980; Forster, 1981; Gough, Alford, & Holley-Wilcox, 1981; Morris, 2006; Stanovich & West, 1981, 1983). Prediction at that time was thought to be too prone to failure and perhaps too costly to be of use to the language comprehension system. Such concerns have, however, been somewhat neglected in this era of renewed interest in prediction. In this article, we return to questions about the potential costs of prediction failure and find that they warrant further empirical investigation.

Prediction in the 20th Century

Prediction is not a new concept in sentence processing. Early theories, especially those on reading, incorporated mechanisms that relied heavily on context and expectation to drive the comprehension process. In an early information-theoretic approach to comprehension reminiscent of recent Bayesian models of language comprehension (Levy, 2008; Kuperberg & Jaeger, 2016), Smith (1971; see also Smith & Holmes, 1971) approached prediction in comprehension as "the reduction of uncertainty" of the comprehender about the meaning of a utterance by "eliminate[ing] some or all of the alternative meanings" (pg. 185-6), a view more recently refined to say, "Prediction is the prior elimination of unlikely alternatives" (Smith, 2004: 25). This viewpoint furthermore suggested that during comprehension readers may not even "extract all the meaning they might acquire if they were to identify every word individually" (pg. 195), a point taken up by Goodman (1967) in his response to the seemly impossible task of the reading process to rapidly and precisely extract the fine-grained detail of text. Goodman argued that reading must

therefore be a selective process that makes use of partial information drawn from perception, a possible precursor for recent proposals of shallow or "good-enough" language processing (Ferreira & Patson, 2007; Levy, 2008; Sanford & Sturt, 2002), combined with an ability to anticipate upcoming information. Evidence from word misrecognition during reading supported these conclusions. Errors in reading aloud tasks were found to more likely reflect a similarity of word meaning than word form, with participants providing a highly expected word in place of the actual word in the input (Goodman, 1965, 1969; Kolers, 1970; Weber, 1968, 1970). Such misrecognition was taken to demonstrate the prioritization of context in a comprehension process that was too fast for individual word identification, with the input merely acting to confirm the comprehender's prior expectations.

These early models of prediction during sentence processing were challenged from both conceptual and experimental perspectives. Conceptually, prediction was thought to be too prone to failure and perhaps too costly to be of use to the language comprehension system. Evidence using Taylor's (1953) cloze task found that highly predictable content words were rare in normal discourse (Bormuth, 1966; Finn, 1977; Gough, 1983; Perfetti, Goldman, & Hogaboam, 1979; Rubenstein & Aborn, 1958; Shanahan, Kamil, & Tobin, 1982; see Luke & Christianson, 2016, for a more recent and extensive investigation), suggesting that a comprehension system that strongly relied on predictive mechanisms would too often receive evidence that was too weak to support a firm prediction. Worse, when prediction seemed possible, the actual input encountered would more often than not be contrary to expectations. Given these results, a highly predictive comprehension system might fail quite frequently, and the predicted

costs of such frequent prediction failures seemed problematic for theories of prediction (Gough, Alford, & Holley-Wilcox, 1981).

Evidence for the existence of such costs, however, was limited. Across several studies that compared word naming latencies in congruent and incongruent contexts to neutral baselines (e.g. for the word "snow", congruent: "The skier was buried in the…"; incongruent: "The bodyguard drove the…"; neutral: "They said it was the…"), Stanovich and West (1981, 1983) found robust evidence for facilitation in congruent sentence contexts but unreliable and limited evidence for the predicted cost of incongruent sentence contexts (Stanovich & West 1981, averaged over 3 studies: facilitation: 56.3 msec, inhibition: -10.0 msec; Stanovich & West 1983, averaged over 11 studies: facilitation: 58.2 msec, inhibition: -14.9 msec). The predicted inflation in error rates for the incongruent condition over the baseline or congruent conditions was also not found. If anything, the trend was for congruent contexts to lead to more errors, contrary to what would be expected given a predictive comprehension mechanism that should have facilitated correct responses (Stanovich & West 1981, averaged over 3 studies: congruent: 1.63%, incongruent: 0.51%; Stanovich & West 1983, averaged over 11 studies: congruent: 1.15%, incongruent: -0.04%). Thus, although early theories of prediction predicted costs for incongruent words, such costs were not empirically born out. Taken together with other findings, the weight of evidence and argument against predictive mechanisms pushed the field away from theories of comprehension as a predictive process (Forster, 1981; Frisson, Rayner, & Pickering, 2005; Schwanenflugel & Lacount, 1988; Schwanenflugel & Shoben, 1985; Traxler & Foss, 2000).

Prediction in the 21st Century

These early empirical failures and conceptual reservations notwithstanding, sentence context is now widely assumed to trigger predictions for upcoming words. The most convincing evidence for this came from a series of studies that manipulated morphosyntactic or morphophonological agreement of articles or adjectives that occurred before an expected word to test if the comprehension system had access to this expected word before it was given in the input (DeLong, Urbach, & Kutas, 2005; Otten, Nieuwland, & van Berkum, 2007; van Berkum, et al., 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003). Wicha, Moreno, and Kutas (2004) presented participants with short stories in Spanish. Each story contained a critical sentence such that a particular noun became highly expected. For example, in the Spanish equivalent of "The story of Excalibur says that the young King Arthur removed from a large stone a…", participants expect the noun "sword". Unlike English, the determiner preceding this expected noun must match in gender with the expected noun. Wicha and colleagues manipulated the morphosyntactic gender of articles so that they were either congruent or incongruent with the expected noun's gender. They found that incongruent articles elicited a frontal positivity between 500 and 700 msec. They argued that this response could only be due to the expectation of the particular idiosyncratic gender of the upcoming noun, and therefore the comprehender had access to the lexical information of this noun prior to its occurrence. Similar results were reported in DeLong, Urbach, and Kutas (2005). Taking advantage of the different morphophonological forms of the indefinite article in English, they presented

participants with sentences like “The day was breezy so the boy went outside to fly…” and manipulated the article before the expected noun to be either congruent “a” or incongruent “an” with the expected noun “kite”. They found that incongruent articles elicited a more negative N400 response compared to congruent articles (though see Nieuwland, et al, 2018, which failed to find this article incongruence effect across a much larger set of participants and laboratories). Given that the different forms of the indefinite article have no semantic import, they argued that the comprehender had predicted the expected noun and had access to its phonological form.

Together with other studies, these experiments provide unambiguous support for predictive mechanisms in language comprehension. However, the consequences such predictions can have on further processing remain unclear. While successful predictions are argued to have a facilitatory effect (Federmeier, 2007), failed predictions could generate processing costs that would limit the overall usefulness of prediction, especially if such costs affected subsequent processing. Indirect evidence for prediction-related processing costs have been reported in many ERP studies as late positive components elicited by unexpected words in high cloze contexts (DeLong, et al, 2011; DeLong, Quante, & Kutas, 2014; Federmeier, et al, 2007; Otten & van Berkum, 2008). These components suggest that certain processes come online to aid in the recovery from failed predictions, but understanding what those processes are remains challenging. They could reflect simple disruptions in processing due to prediction failure and the processing of an error signal (Van Petten & Luka, 2012), or might reflect particular recovery processes such as the inhibition of failed predictions (Kutas, 1993), revision of high-level sentence and discourse representations (Brothers,

Swaab, & Traxler, 2015), or adaptation of expectations for future predictive use (Kuperberg & Jaeger, 2015).

Whether these costs for prediction failure more directly impact ongoing language processing, however, has not been directly addressed. In this study, we explore one possible consequence that failed predictions could have on the ongoing processing of semantic context. If the comprehender becomes less certain of or less reliant on their predictive mechanisms due to a recent failure, they may fail to reap the benefits of helpful semantic context that would otherwise be used to aid in processing (Lau, Holcomb, & Kuperberg, 2013). Thus it is important to establish the speed with which the comprehender recovers from prediction failures. Considering the range of possibilities, at one end, recovery might be very rapid, leading to no disruption in the comprehender's subsequent use of semantic context. The comprehender might ignore or immediately discard a failed prediction, rapidly returning to business as usual. Such a finding would militate against earlier concerns of researchers on the cumulative costs of prediction failure because any cost would be quickly overcome and unable to accumulate to impact ongoing language processing. Empirically, support for this position may be seen in the lack of evidence for a processing cost of prediction failure across a variety of methodologies, including the naming times studies of Stanovich & West (1981, 1983) noted above. In ERPs for instance, the N400 itself appears to be insensitive to failed predictions (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kutas & Hillyard, 1984; Lau, Almeida, Hines, & Poeppel, 2009). As summarized in Van Petten & Luka (2012), instead of reflecting a response to incongruence, the large negative amplitude of the N400 should be seen as a default response to words that is

reduced in the presence of supportive semantic context. Reading times measured in eye-tracking also appear to be insensitive to cost of prediction failure (Ehrlich & Rayner, 1981; Luke & Christianson, 2016; Staub, 2015). Use of other techniques, such as cumulative semantic interference, finds no additional cost to name semantically same-category pictures after completion of a high cloze sentence compared to basic picture naming trials (Kleinman, Runnqvist, & Ferreira, 2015). This suggests that prediction failures themselves may be relatively costless to the system, perhaps reflecting a gradual passive bottom-up pre-activation process instead of a fully-fledged tokening of a particular lexical item into working memory (DeLong, Troyer, & Kutas 2014; Kuperberg & Jaeger, 2015).

Alternatively, recovery might be more protracted, with prediction failures actively disrupting the processing system temporarily and perhaps even accumulating to affect overall global processing. While early evidence for a slowdown on word recognition itself was limited, there was a small but consistent effect across studies (Stanovich & West, 1981, 1983; see also Fischler & Bloom, 1979; Forster, 1981; Gough, Alfrod, & Holley-Wilcox, 1981; Schuberth & Eimas, 1977; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985). The idea that such costs might accumulate over time was also recently supported by Lau, Holcomb, and Kuperberg (2013). In a semantic priming paradigm, they found that the proportion of related vs. unrelated prime-target pairings affected the N400 response. In their study, they varied the proportion of related prime-target pairs within two blocks to be either 10% or 50% of the items, but held the semantic association between prime and target constant. Target words in both blocks elicited a standard N400 effect; however, this effect

was highly attenuated in the low-proportion block compared to the high proportion block. This suggests that the accumulation of prediction failures leads the system to stop predicting even though supportive semantic context is available, though it is unclear whether this attenuation emerges immediately, or whether it results from the accumulation of costs across a block of trials. These studies suggest that prediction failures are disruptive and that these disruptions can affect subsequent processing, perhaps reflecting a more active item-specific prediction.

At issue then is whether recovery from prediction failure is rapid or protracted. To investigate this, we manipulated the congruence of a preceding article and the predictiveness of a local adjective in high cloze Italian sentences. We used the form of Italian articles to cue the comprehension system to the status of a prediction; article gender that was incongruent with the expected noun's gender acted as an early cue to the system of a prediction failure for the upcoming noun. We also inserted an adjective between the article and the noun that was either predictive or neutral with respect to the actual upcoming noun. Such local adjective-noun pairs are known to elicit prediction-like behavior. In addition to Lau, Holcomb, and Kuperberg (2013) above, Fruchter, Linzen, Westerlund, and Marantz (2015) investigated lexical preactivation of a noun driven by a preceding adjective using magnetoencephlography. They found a reduction of activity in the left medial temporal gyrus to nouns preceded by an adjective that was predictive compared to those that were not.

The two theoretical possibilities make different experimental predictions. If recovery from prediction failure is a rapid process, we expect comprehenders to make immediate use of local semantic context provided by the adjective to

preactivate the noun regardless of whether the system was cued to a prediction failure or not. Such preactivation should reduce the N400 response to nouns in locally predictive contexts compared to neutral contexts by roughly equal measures. If recovery from prediction failure is a more protracted affair, we expect comprehenders will be unable to make immediate use of the local semantic context provided by the adjective to preactivate the noun when the system has been cued to a prediction failure. The lack of preactivation should lead to similar N400 amplitudes in both locally predictive and neutral contexts when a prediction failure has recently been signaled.

Experiment

Materials and methods

Participants. 30 native Italian speakers (14 female, average age 28) from the University of Oxford and surrounding community participated in this study for £20 each.

Materials. We manipulated 40 sentences in Italian with high cloze noun completions in a 2 (article gender *congruence* with the global context) x 2 (local adjective's *predictiveness* of the noun) design, as shown in Table 1. Sentence contexts were constructed to elicit a noun phrase and were normed using a cloze procedure. 198 native Italian speakers were asked to complete each sentence context fragment. Cloze probability was calculated for each context as the proportion of speakers choosing to complete that context with a particular noun, yielding 20 contexts with high cloze feminine nouns and 20 contexts with high cloze masculine nouns. The average cloze probability over all 40 sentence contexts was 0.76 (min: 0.31; max: 0.98).

Following research manipulating gender agreement as an early cue for prediction failure (Otten & Van Berkum, 2008; Van Berkum, et al., 2005; Wicha, et al., 2004), we manipulated article gender to be either *congruent* or *incongruent* with the expected noun's gender such that an incongruent article unambiguously cued a global prediction failure. Noun phrases requiring the opposite gender of our high cloze sentence contexts were paired together and swapped with one another such that the gender mismatching noun itself was now low cloze given the global sentence context (average incongruent cloze: 0%). Note that no sentences were, strictly speaking, ungrammatical. The gender of the article was always grammatically appropriate to the noun in our stimuli. The manipulation was only whether that noun was expected in the context, such that the gender of unexpected nouns could act as an early cue to prediction failure.

To manipulate local semantic context, an adjective was inserted between the article and noun so that the adjective was either *predictive* or *neutral* with respect to the upcoming noun. Following Lau, Holcomb, and Kuperberg (2013), adjective-noun pairings were selected based on co-occurrence frequencies retrieved from the "La Repubblica" corpus of written Italian (Baroni, Bernardini, Comastri, Piccioni, Volpi, Aston & Mazzoleni, 2004). Predictive adjectives were highly predictive of the upcoming noun (average Pr(noun | adjective) = .55; min Pr(noun | adjective) = .33; max Pr(noun | adjective) = .96). Neutral adjectives were not strongly predictive of either the upcoming noun (average Pr(noun | adjective) = .003) or any other noun (average of max Pr(noun | adjective) = .06). Cloze probabilities for the resulting sentence contexts with adjectives were obtained from 80 new Italian participants in a cloze procedure task. Average

cloze probabilities were calculated for each condition (congruent predictive: 0.80, congruent neutral: 0.64, incongruent predictive: 0.10, incongruent neutral: 0.00). A binomial linear effects model found main effects of both the congruence of the global context (Est. = 0.335, t = 22.879, $p < .001$) and the predictiveness of the local adjective (Est. = 0.063, t = 6.446, $p < .001$), but no interaction between the two factors (Est. = 0.015, t = 1.483, $p = .141$). While we are cautious in our interpretation of this result given possible floor effects due to the extremely low cloze probability of the incongruent neutral, the results suggest that the addition of the adjective had similar additive effects on cloze probabilities across both congruent and incongruent conditions.

Examples of the final sentence stimuli are given in Table 1. These sentences were counterbalanced across four lists such that every participant saw 10 sentence stimuli per condition. An additional 200 filler sentences were included, 120 of which examined the processing of auxiliaries after animate and inanimate subjects while the other 80 masked the local adjective predictiveness manipulation and presented participants with a more diverse and natural set of sentence constructions. All fillers were grammatical. A comprehension question was asked after each sentence.

TABLE 1 ABOUT HERE

Procedure. Participants were tested in a single session in a soundproof, electrically shielded room. They were seated in a chair in front of a 32" HD LED screen (Samsung Smart TV) positioned approximately 120 cm away and instructed to read the sentences for comprehension while avoiding eye and body movements and blinks. The session began with a short set of practice sentences

before presentation of the experimental stimuli to accustom participants to the stimulus presentation.

Sentences were presented one word at a time in the center of the screen in black 50-point serif typeface, on a light grey background. Each trial was initiated by a fixation cross that remained for 2 sec. Sentence stimuli were then presented using rapid serial visual presentation. Each word remained on the screen for 200 msec and was followed by a 300 msec blank screen for a stimulus onset asynchrony of 500 msec. A comprehension question appeared on the screen 1000 msec after the end of each sentence. Participants had to answer it by pressing the appropriate button on a computer mouse.

Electrophysiological recording. EEG was recorded on a 64-channel ANT Neuro system, mounted in an elastic cap, and referenced to the Cz electrode. Blinks and eye movements were registered by placing an electrode under each eye. Electrode impedance was kept below 20 kΩ throughout the experiment. The EEG was amplified with an ANT Neuro amplifier and sampled with a frequency of 512 Hz.

Data Analysis. Offline preprocessing and measurement of EEG data was done in Matlab using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). Artifact detection/correction was done using algorithms from FASTER (Nolan, Whelan, & Reilly, 2010). Channels with local artifacts were interpolated when possible. EEG data was filtered (0.1-40 Hz), segmented -200 to 1000 msec time-locked to the onset of the target noun, rereferenced to the average of all channels, and baseline corrected using the -200-0 msec time window prior to the target noun onset. Subject averaged ERPs were formed from trials free of ocular and muscular artifacts. Seven participants were eliminated

due to excess artifacts leaving them with fewer than 65% of the total trials. Grand average ERPs were formed using the remaining 23 participants. The final trial count average (and standard deviation) by condition per participant was Congruent-Predictive, 8.96 (1.33); Congruent-Neutral, 8.91 (1.16); Incongruent-Predictive, 9.17 (0.89); and Incongruent-Neutral, 9.00 (1.17), resulting in 3128 total trials for analysis (congruent predictive: 768, congruent neutral: 776, incongruent predictive: 792, incongruent neutral: 792).

Visual inspection of the grand average ERPs revealed two time windows of interest: 250-500 msec, reflecting the N400, and 500-1000 msec, reflecting a post-N400 component. Post-N400 components have been reported in several similar studies investigating high cloze sentence contexts, with nouns in incongruent conditions typically eliciting frontal positivities when compared to congruent nouns (DeLong, et al, 2011; DeLong, Quante, & Kutas, 2014; Federmeier, et al, 2007; Otten & van Berkum, 2008; and see Van Petten & Luka, 2012, Table 2 for a wider survey of the literature). Assessment of amplitude differences within these two time windows was conducted using the lme4 (v1.1-9) and lmerTest (v2.0-29) packages in R (R Development Core Team, 2010). Linear mixed effects models with random by-subject and by-item intercepts and slopes were constructed on mean ERP amplitudes between 250-500 msec and 500-1000 msec post stimulus onset over a subset of 52 electrodes divided into two levels of Hemisphere (left/right) and Anteriority (anterior/posterior), defining four quadrants (left anterior: Fp1, AF3, AF7, F1, F5, F9, FC1, FC3, FC5, FT7, and FT9; right anterior: Fp2, AF4, AF8, F2, F6, F10, FC2, FC4, FC6, FT8, and FT10; left posterior: C3, C5, CP1, CP3, CP5, P1, P3, P7, P9, T7, TP7, TP9, PO1, O1, and O9; right posterior: C4, C6, CP2, CP4, CP6, P2, P4, P8, P10, T8, TP8, TP10,

PO2, O2, and O10) with Congruence and Predictiveness included as experimental factors. All factors were sum-coded to allow for ANOVA-style analysis. Model means and 95% confidence intervals in bar plots were calculated using the effects (v3.0-4) package.

Results

Comprehension accuracy. Average response accuracy to the comprehension questions was very high at 91%.

ERPs on target noun.

N400 results. Figures 1 and 2 illustrate the N400 response to target nouns that were neutral or predicted by a preceding adjective under conditions where the article's gender was congruent with and incongruent with the expected noun given the sentence context. A by-subject and item linear model in the 250-500 msec time window across four quadrants revealed a significant main effect of Anteriority (Est. = -0.245, SE = 0.055, t = -4.442, p < .001) and Hemisphere (Est. = 0.180, SE = 0.055, t = 3.265, p = .001) and a significant two-way interaction between Congruence and Anteriority (Est. = -0.396, SE = 0.055, t = -7.198, p < .001), Predictiveness and Anteriority (Est. = -0.1295, SE = 0.055, t = -2.350, p = .019) and Anteriority and Hemisphere (Est. = 0.151, SE = 0.055, t = 2.732, p = .006). A significant three-way interaction was found between Congruence, Predictiveness, and Anteriority (Est. = -0.173, SE = 0.055, t = -3.147, p = .002). The four-way interaction was not significant (p = .507). Table 2 reports the linear model estimates, standard errors, t values, and p values. Figure 3 illustrates the overall quadrant analysis.

FIGURES 1, 2, AND 3 AND TABLE 2 ABOUT HERE

The three-way interaction of our experimental factors with Anteriority was driven by nouns with predictive adjectives in congruent contexts eliciting greater positivity over posterior regions (congruent predictive: 0.959 μV, congruent neutral: 0.289 μV, incongruent predictive: -0.252 μV, incongruent neutral: -0.163 μV) and greater negativity over anterior regions (congruent predictive: -0.929 μV, congruent neutral: -0.387 μV, incongruent predictive: 0.140 μV, incongruent neutral: 0.053 μV) when compared to nouns in the other three conditions, shown in Figure 3. Model contrasts over the anterior and posterior regions demonstrated that the effect of predictiveness (predictive vs. neutral) was significant in the congruent conditions (anterior: Est. = -0.271, SE = 0.111, t = -2.438, p = .015; posterior: Est. = 0.335, SE = 0.111, t = 3.014, p = .003) but not in the incongruent conditions (anterior: Est. = 0.043, SE = 0.110, t = 0.396, p = .692; posterior: Est. = -0.044, SE = 0.110, t = -0.404, p = .686).

Post-N400 results. Figure 4 illustrates the post-N400 response to target nouns that were neutral or predicted by a preceding adjective under conditions where the article's gender was congruent with and incongruent with the expected noun given the sentence context. A by-subject and item linear model in the 500-1000 msec time window across four quadrants revealed a significant main effect of Hemisphere (Est. = 0.133, SE = 0.062, t = 2.141, p = .032) and a significant two-way interaction between Congruence and Anteriority (Est. = -0.372, SE = 0.062, t = -5.974, p < .001) and a significant three-way interaction between Congruence, Predictiveness, and Anteriority (Est. = -0.157, SE = 0.062, t = -2.516, p = .012). The four-way interaction was not significant (p = .287). Table 3 reports the linear model estimates, standard errors, t values, and p values. Figure 5 illustrates the overall quadrant analysis.

FIGURES 4 AND 5 AND TABLE 3 ABOUT HERE

The three-way interaction of our experimental factors with Anteriority was driven by nouns in incongruent contexts eliciting a greater positivity over anterior regions (congruent predictive: -0.416 μV, congruent neutral: -0.306 μV, incongruent predictive: 0.576 μV, incongruent neutral: 0.105 μV) and a greater negativity over posterior regions (congruent predictive: 0.455 μV, congruent neutral: 0.292 μV, incongruent predictive: -0.666 μV, incongruent neutral: -0.157 μV) when compared to nouns in the congruent condition, shown in Figure 5. Model contrasts over the anterior and posterior regions demonstrated that the effect of congruence was significant in the anterior regions (Est. = -0.701, SE = 0.176, t = -3.988, p < .001) and posterior regions (Est. = 0.785, SE = 0.176, t = 4.461, p < .001). In the incongruent condition, there was a visual trend for nouns to elicit greater anterior positivity and posterior negativity in the predictive adjective condition compared with the neutral adjective condition within this time window, but this differences only reached significance in the posterior region (anterior: Est. = 0.235, SE = 0.124, t = 1.905, p = .057; posterior: Est. = -0.254, SE = 0.124, t = -2.057, p = .040).

To further explore the apparent effect of predictiveness in the incongruent condition, we analyzed a more focused time window from 650-800 msec. A by-subject and item linear model in the 650-800 msec time window across four quadrants revealed a significant main effect of Anteriority (Est. = 0.198, SE = 0.070, t = 2.809, p = .005) and Hemisphere (Est. = 0.229, SE = 0.070, t = 3.262, p = .001) and a significant two-way interaction between Congruence and Anteriority (Est. = -0.327, SE = 0.070, t = -4.650, p < .001). A significant three-way interaction was found between Congruence, Predictiveness, and Anteriority

(Est. = -0.230, SE = 0.070, *t* = -3.271, *p* = .001). The four-way interaction was not significant (*p* = .348). Table 4 reports the linear model estimates, standard errors, *t* values, and *p* values. Figure 6 illustrates the overall quadrant analysis.

FIGURE 6 AND TABLE 4 ABOUT HERE

In addition to the effect of congruence seen in the broader 500-1000 msec time window analysis, analysis of the more focused 650-800 msec time window revealed that nouns in the incongruent-predictive condition elicited a greater positivity over anterior regions (congruent predictive: -0.199 μV, congruent neutral: -0.018 μV, incongruent predictive: 0.796 μV, incongruent neutral: 0.184 μV) and greater negativity over posterior regions (congruent predictive: 0.287 μV, congruent neutral: -0.014 μV, incongruent predictive: -0.947 μV, incongruent neutral: -0.172 μV) when compared to nouns in the incongruent-neutral condition, as shown in Figure 6. Model contrasts comparing incongruent conditions to congruent conditions over the anterior and posterior regions again demonstrated a significant positivity over anterior regions (Est. = -0.598, SE = 0.199, *t* = -3.004, *p* = .002) and a significant negativity over posterior regions (Est. = 0.711, SE = 0.199, *t* = 3.567, *p* < .001). Within the incongruent condition, nouns in the predictive condition elicited a significantly greater negativity over anterior regions (Est. = 0.306, SE = 0.140, *t* = 2.189, *p* = .029), and a significant positivity over posterior regions (Est. = -0.387, SE = 0.140, *t* = -2.768, *p* = .006). No significant effects for predictiveness were revealed within the congruent condition (anterior: *p* = .523; posterior: *p* = .336).

ERPs on preceding article.

Figure 7 illustrates an emerging positivity 300 msec after article onset with gender incongruent articles eliciting a more positive ERP than gender

congruent articles. A by-subject and item linear model in the 300-500 msec time window across four quadrants revealed a significant main effect of Hemisphere (Est. = 0.312, SE = 0.056, t = 5.565, p < .001) and a significant two-way interaction between Congruence and Anteriority (Est. = -0.171, SE = 0.056, t = -3.048, p = .002) and Anteriority and Hemisphere (Est. = 0.137, SE = 0.056, t = 2.470, p = .014). Table 5 reports the linear model estimates, standard errors, *t* values, and *p* values.

FIGURE 7 AND TABLE 5 ABOUT HERE

Discussion

Prediction has come to play a central role in our understanding of the mechanisms of language comprehension. In spite of early concerns about the robustness of a predictive language comprehension architecture, recent evidence in support of predictive mechanisms has been well established, particularly in studies examining the processing of agreement forms that depend on an upcoming expected item (DeLong, Urbach, & Kutas, 2005; Otten & Van Berkum, 2008, 2009; Otten, et al., 2007; Van Berkum, et al., 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003; Wicha, et al., 2004; but cf. Nieuwland, et al, 2018). These findings have supported a view of comprehension as an active process that can predict expected lexical items, especially in sentence contexts with high constraint. As a result, we may begin to turn away from questions concerning whether language comprehension is predictive to those addressing how such predictive mechanisms operate (Kutas, DeLong, & Smith, 2011). Questions about the potential costs of prediction failure are of particular interest, especially as these questions proved empirically intractable

for early theories of prediction (Fischler & Bloom, 1980; Forster, 1981; Gough, Alford, & Holley-Wilcox, 1981; Morris, 2006; Stanovich & West, 1981, 1983).

In this study, we aimed to address two related questions about the operation of predictive comprehension: what are the consequence of prediction failure on subsequent processing and what do those consequences tell us about how a comprehender recovers from prediction failures? We used the gender of a preceding article to cue the comprehender to the upcoming prediction failure of an expected noun while manipulating the local semantic context between the adjective and target noun to be predictive or neutral. Our results showed that prediction failures cued by incongruent articles blocked the use of the local semantic context given by the adjective on subsequent processing of the noun as measured by the N400 response to that target noun, in spite of the similar increase in cloze probability provided by the predictive adjective in congruent and incongruent conditions in offline cloze probability measures. This result suggests that recovery from prediction failure is not a rapid costless process. Instead, prediction failure appears to limit the potential processing advantage provided by local semantic context. In this study, prediction failure prevented the processing of a target noun from receiving the facilitation that would have otherwise been expected if the system had used the local semantic context provided by the adjective to preactivate the noun. We also found that, after the N400, target nouns in incongruent conditions elicited a frontal positivity similar to those reported in other studies (DeLong, Urbach, Groppe, & Kutas, 2011; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). In our study, this post-N400 component was also sensitive to local semantic context, yielding a more sustained effect when the local semantic context was predictive. We first discuss

the results related to the N400, and then turn to those related to the post-N400 component.

The lack of a reduction in the amplitude of the N400 for predictive local semantic context after a prediction failure reinforces earlier concerns about the general costs of prediction failure (Gough, Alford, & Holley-Wilcox, 1981) and raises questions of just how frequent such failures may be. In a recent and extensive study of sentential constraint and word predictability, Luke and Christianson (2016) provide an answer to this question. They measured cloze probabilities for every word in 55 every day text passages, including online news articles, popular science magazines, and works of fiction. They found on average that the actual target content word was the most frequent response for only 21% of all content words, meaning that, for about four fifths of the content words, some word other than the target word was more expected by participants. Focusing just on nouns themselves did not greatly improve the picture. The actual target noun was the most frequent response only 26% of the time, suggesting that there was a more expected noun for about three quarters of the nouns in these texts. Such a high rate for potential prediction failure in combination with our results that prediction failure disrupts at least some aspects of subsequent processing raises important questions about the robustness of prediction during language comprehension. Part of the solution to these questions concerns the underlying processes at play during recovery from prediction failures.

One possibility is that prediction failure leads the system to globally down regulate its use of prediction. An overall reduction in predictive processing could explain why Lau, Holcomb, and Kuperberg (2013) found diminished reduction of

the N400 on target trials in low relatedness blocks which discouraged predictive processing compared to their high relatedness blocks. In a low relatedness block a comprehender would encounter a high rate of prediction failure, leading them to globally reduce their reliance on predictive processes. However, in our study comprehenders did not seem to simply reduce their reliance on prediction by some amount that accumulated over a number of trials and reduced their reliance on prediction overall. If comprehenders had simply lowered their global reliance on predictive processes enough to suspend predictive processing by the magnitude seen in the incongruent condition after a prediction failure, such a suspension should have also affected the congruent conditions to the same degree since either condition could have followed, contrary to the results. While the N400 showed little sign of reduction in the incongruent condition, our congruent condition continued to show robust reduction of the N400 in locally predictive contexts. Given that these trials were intermixed, this suggests that local semantic context was temporarily disrupted on more of a trial-by-trial basis. Additionally, a large change of the global rate of predictive processing on a trial-by-trial basis would raise questions concerning how the comprehender would ever recover such that prediction would be possible on the trials where prediction could be successful within the context of the study. By globally reducing their reliance on predictive processes, comprehenders would not be predicting on those trials where prediction would succeed and thus would not have distinguished congruent from incongruent trials, treating them all the same. Thus the cue that prediction was successful in congruent trials would have been missed. Of course, Lau, Holcomb, and Kuperberg (2013) demonstrate that comprehenders can track the accumulation of prediction failures and use this to

ultimately reduce their global reliance on predictive processing, perhaps in a more incremental fashion, but such a global mechanism seems unlikely to explain our results.

This suggests that recovery from prediction failures in our study reflected more local processing decisions, albeit ones that could lead to global changes given the global frequency of their occurrence. These local processing decisions may have resulted in processing resources being temporarily diverted away from predictive mechanisms as the system recovered. Given the nature of the cue to prediction failure in our study, there are several processes that the comprehender might have engaged in while recovering from prediction failure.

Because predictive mechanisms are thought to preactivate and precompute representations, comprehenders might engage in processes required to discard their prediction, either through some active inhibition process or by rapid decay. Early on, Kutas (1993), following arguments by Halgren (1990), suggested that such inhibitory processes might be necessary to interpret unexpected but congruent targets. Federmeier, et al., (2007) also considered the relevance of inhibition in an account of the late positivities elicited by unexpected but congruent words in high-cloze compared to low-cloze contexts, suggesting that such a process may be necessary to override the narrow scope of facilitation associated with high-cloze contexts (Schwanenflugel & La Count, 1988; see also, DeLong, et al., 2011, and Thornhill & Van Petten, 2012). Given that the cue to prediction failure and the local semantic context in our study occurred prior to the target word, comprehenders could have been engaged in processes that were inhibiting the expected target noun in preparation for an unexpected but congruent alternative.

Research has also suggested that comprehenders may be engaged in revision of their sentential and discourse representations. Brothers, Swaab, and Traxler (2015) found that the amplitude of the late positivity elicited by unexpected but congruent target words tracked the plausibility of the actual target word given its sentential context. They proposed that the less plausible an unexpected target is the more integrative processing is needed to successfully build the correct representation. Incongruent article gender in our study could have invited comprehenders to begin revising their sentential representations prior to receiving the target noun, directing resources away from processing the local semantic context.

Prediction failure has also been seen as an opportunity for comprehenders to learn by adapting their representations for future predictive use (Kutas, DeLong, & Smith, 2011). Kuperberg and Jaeger (2015) suggest that the costs observed for prediction failure may reflect processes by which comprehenders update their representations to better reflect the structure of the environment they are in. Such an adaptive process would adjust the relationship between the comprehender's prior expectations and the input they received to generate better future predictions. Boudewyn, Long, and Swaab (2015) found that comprehenders can rapidly adapt their predictions during sentence processing, weakening their expectations for high-cloze nouns when given semantically inconsistent input. Comprehenders in our study may have failed to reap the benefits of supportive local semantic context because their processing system was occupied with updating their representations once cued by incongruent article gender to the failure of their prediction.

Any of these responses to prediction failure could have diverted processing resources away from predictive processing, thus limiting the use of local semantic context without disrupting the system's overall global behavior. Importantly, these processes are not incompatible with one another and could coincide during the recovery process. Processes that suppress failed expectations could operate in tandem with those that revise higher-level sentence and discourse representations. The strengthening and weakening associations to improve the outcome of future predictions could also proceed as particular lexical and sentential revisions are being made.

Regardless of whether the blocking of local semantic context arises from any of these local recovery mechanisms or from global processing considerations, we find that recovery from prediction failure is a protracted process. How protracted the recovery process is, however, is still unclear. One possible source of evidence may be the late positivities observed in response to unexpected but plausible target words. DeLong, et al., (2011), for instance, suggests that onset of late positivities may be linked to whatever processes are brought online to recover from prediction failures. Their finding of an earlier onset of their late positivity compared to Federmeier, et al., (2007) could be related to the early cue delivered by an incongruent article form prior to receiving the target noun. Comprehenders using that information in combination with the input of an unexpected noun may have been able to initiate recovery processes earlier than they otherwise could be initiated, perhaps shortening recovery time. While the availability of early cues may help comprehenders to more rapidly initiate a recovery process, a host of other factors are likely to play a role in the duration of recovery depending on what such recovery requires. The

strength of a prediction, related to the amount of constraint delivered by the context, is likely to be one significant factor, though others, including relatedness of pre-activated representations, the strength of competing expectations, the time between pre-activation and the cue to prediction success or failure, and the difficulty of input-driven lexical access all seem to be likely candidates for future research.

Though not explicitly predicted given the main focus of our study, we found that target nouns following incongruent articles elicited a post-N400 component with a scalp distribution similar to the late frontal positivities found in other studies (DeLong, et al., 2011; Federmeier, et al., 2007). Although our component had both a frontal positivity and a posterior negativity, we will continue to use the term late frontal positivity in keeping with this emerging literature.

Late frontal positivities are thought to arise when the target of a prediction is merely unexpected instead of anomalous. In a recent study, DeLong, Quante, and Kutas (2014) directly compared unexpected and anomalous continuations to expected continuations in high cloze sentence contexts. They found that unexpected continuations elicited late frontal positivities whereas anomalous continuations elicited late posterior positivities. While we did not explicitly manipulate our target nouns in the incongruent condition to be unexpected or anomalous, this dissociation suggests that the items in our study may have been unexpected but plausible for our participants. However, a review of our items indicated that only 20 of the 80 incongruent sentences were unexpected but plausible. The other 60 incongruent sentences were more implausible or anomalous, though intuition suggests that this was a graded

distinction. This suggests that late frontal positivities are not always associated with an unexpected but plausible target continuation. Because our study manipulated prediction failure with an early cue of gender incongruence of the article with the most expected noun, and not the target noun itself as it was in DeLong, Quante, and Kutas (2014), the late frontal positivity we observe may have been driven by different factors.

One possibility is that the late frontal positivity we observed was related to early stages of the recovery process made available by the early article gender cue in our study. As the semantic context of a sentence unfolds, a variety of semantically possible nouns may have been preactivated prior to the article. Although the gender of the most expected noun failed to agree with the article's gender in the incongruent conditions, some semantically unexpected but plausible nouns would have still been compatible with the global semantic context. Comprehenders trying to maintain global coherence with the semantic context they had been given may have used the gender incongruence cue to update their predictions for unexpected but plausible nouns. Having these unexpected but plausible nouns active at the target noun may have elicited a late frontal positivity, though future research will be needed to explore this possibility.

Interestingly, the late frontal positivity found in our study was affected by the predictiveness of local semantic context. While target nouns in the incongruent condition generally elicited a late positivity, this effect was more robust and sustained when the local semantic context was predictive of the target noun. No post-N400 effect of predictiveness was found in the congruent conditions where prediction was successful. This suggests that locally predictive

semantic context generated additional sustained costs when comprehenders are trying to recover from prediction failure in our study; a surprising finding given that predictive local semantic context was provided to help aid subsequent processing. The nature of these costs is unclear; however, several options seem possible.

One option is that our local semantic contexts continued to support the prediction of the expected noun even though such a prediction would ultimately fail because of a gender agreement violation. If recovery from prediction failure involves processes that suppress or discard the failed prediction, then information supporting that prediction could have led to interference in the recovery process, indexed by a more robust late frontal positivity. This would suggest that, while the comprehender cannot use local semantic context after prediction failure to preactivate a new lexical item, they are unable to prevent the predictiveness of local semantic context to impact their revision process. To address this possibility, we analyzed the predictiveness of our adjectives for the expected noun in incongruent sentences, using co-occurrence frequencies of the correct gender matched adjective form and expected noun retrieved from the "La Repubblica" corpus (Baroni, et al., 2004). We found that our predictive adjectives were not predictive of expected nouns in incongruent sentences (Pr(exp. N | Adj) = 0.00034), suggesting that the robustness of the late positivity in the predictive condition was not driven by a relationship between the local semantic context and the failed prediction for the expected noun.

Another option related to an inhibitory account of late frontal positivities suggests that the predictions normally licensed by our local predictive semantic contexts triggered further suppression as part of the protracted recovery from

prediction failure. Under this hypothesis, the suppression mechanism reflected in late frontal positivities is not only protracted but also indiscriminate concerning which predictions it is discarding. Although the predictions related to local semantic context were not related to those of global semantic context, it could be that the suppression mechanism triggered by prediction failure is unable to distinguish between these cases, and treats all predictions as suspect during the drawn out process of recovery. Such an account would align nicely with the lack of reduction in the N400 amplitude, suggesting that the reason local predictive semantic context was not uses to recover from prediction failure is because the predictions licensed by the local semantic context are being suppressed alongside global predictions.

A different option from these two above is that the robustness of the late positivity reflects the difficulty of integrating a strongly semantically coherent noun phrase with a globally incompatible meaning. The local predictive context may have served to further highlight the incompatibility of the target noun with the global context, triggering a more robust error signal reflected in a more robust late positivity. This would be compatible with our intuitions concerning our items as more anomalous than merely unexpected, though the question remains why this triggered a late frontal positivity rather than a late posterior positivity given that anomality is thought to elicit posterior positivities after the N400 (DeLong, Quante, & Kutas, 2014). Again, the dynamics of our early cue to prediction failure may have driven the distribution of our post-N400 component in ways that are not yet understood.

Conclusion

Predictive mechanisms have traveled a rocky road in the history of language comprehension research. Early theories gave way to skepticism in the wake of empirical and conceptual problems, and while careful empirical work over the last decade has led to a resurgence of interest, earlier worries surrounding prediction cannot be dismissed out of hand. We have demonstrated that with the benefits of prediction also come possible costs that any theory of predictive mechanisms will need to address if it is to continue to see prediction as a viable core mechanism for language comprehension.

Acknowledgements

We would like to thank the audience of the 29th annual CUNY Human Sentence Processing Conference for initial feedback on our work. This research was made possible by a research support grant from Faculty of Linguistics, Philology, and Phonetics at the University of Oxford and the St. Hugh's College Fellows Research Allowance.

Declaration of Interest Statement

The authors report no conflict of interest.

References

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065-1086.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364-390.

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., Mazzoleni, M. (2004). Introducing the "la Repubblica" corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. *Proceedings of LREC 2004.*

Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly, 1*(3), 79-132.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 607-624.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135-149.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120394.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150-162.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*(12), 631-645.

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203-1207.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-1121.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*(3), 293-321.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641-655.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*, *1146*, 75-84.

Ferreira, F., & Patson, N. D. (2007). The 'good enough'approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71-83.

Finn, P. J. (1977). Word frequency, information theory, and cloze performance: A transfer feature theory of processing in reading. *Reading Research Quarterly, 13*(4), 508-537.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 1-20.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*(4), 465-495.

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say "book"? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience, 30*(6), 768-780.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 862-877.

Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *Journal of Cognitive Neuroscience, 27*(10), 1912-1935.

Goodman, K. S. (1965). A linguistic study of cues and miscues in reading. *Elementary English, 42*, 639-643.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, *6*(4), 126-135.

Goodman, Kenneth S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly, 5*(1), 9-30.

Gough, P. B. (1983). Context, form, and interaction. In K. Rayner (Ed.), *Eye Movements in Reading* (pp. 203-211). New York: Academic Press.

Gough, P. B., Alford, J. A., & Holley-Wilcox, P. (1981) Words and context. In O. J. L. Tzeng & H. Singer (Eds.), *Perception of Print: Reading Research in Experimental Psychology*, (pp. 85-102). Hillsdale, NJ: Erlbaum

Halgren, E. (1990). Insights from evoked potentials into the neuropsychological mechanism of reading. In A. B. Scheibel & A. F. Wechslcr (Eds), *Neurobiology of Higher Cognitive Function* (pp. 103-149). New York: Guilford Press.

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2015). Single-word predictions of upcoming language during comprehension: evidence from the cumulative semantic interference task. *Cognitive Psychology*, *79*, 68-101.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 262-284.

Kolers, P. A. (1970). Three stages of reading. In H. Levin & J. P. Williams (Eds.), *Basic Studies on Reading* (pp. 90-118). New York: Basic Books.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32-59.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533-572.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain: Using our Past to Generate a Future* (pp. 190–207). New York, NY: Oxford University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161-163.

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and Language*, *111*(3), 161-172.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484-502.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*,*106*(3), 1126-1177.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22-60.

Morris, R. K. (2006). Lexical processing and sentence context effects. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 377–401). London, UK: Elsevier.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Meziere, D., Barr, D., Rousselet, G., Ferguson, H., Busch-Moreno, S., Fu, X., Kulakova, E., Tuomainen, J., Husband, E. M., Donaldson, D., Kohút, Z., Rueschemeyer, S.-A., and Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife, 7*:e33468.

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, *192*(1), 152-162.

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC neuroscience*, *8*, 1.

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464-496.

Perfetti, C. A., Goldman, S. A., & Hogaboam, T. W. (1979). Reading skill and the identification of words in discourse context. *Memory & Cognition, 7*(4), 273-282.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105-110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514-528.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504-509.

Rubenstein, H., & Aborn, M. (1958). Learning, prediction, and readability. *Journal of Applied Psychology, 42*(1), 28-32.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382-386.

Schuberth, R. E., & Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 27-36.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 344-354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*(2), 232-252.

Shanahan, T., Kamil, M., & Tobin, A. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 17*(2), 229-255.

Smith, F. (1971). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read*. Holt, Rinehart, Winston, Inc.

Smith, F. (2004). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read* (6th Ed.). Lawrence Erlbaum Associates.

Smith, F. & Holmes, D. L. (1971). The independence of letter, word, and meaning identification in reading. *Reading Research Quarterly, 6*(3), 394-415.

Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 658-672.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*(1), 1-36.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311-327.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297-314.

Taylor, W. L. (1953). ‘‘Cloze procedure": A new tool for measuring readability. Journalism Quarterly, 30(4), 415-453.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382-392.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1266-1282.

Tulving, E. & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology, 66*, 319-327.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443-467.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.

Weber, R. M. (1968). The study of oral reading errors: A survey of the literature. *Reading Research Quarterly, 4*, 96-119.

Weber, R. M. (1970). First graders' use of grammatical context in reading. In H. Levin & J. Williams (Eds.), *Basic Studies in Reading*. New York: Basic Books.

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potential to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165-168.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event-related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex, 39*, 483-508.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration,

gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*, 1272-1288.

Wlotko, E. W., & Federmeier, K. D. (2013). Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in psychology*, *4*, 181.

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics*, *36*(3), 277-284.

Table 1: Example stimulus

	Predictive Adjective	**Neutral Adjective**
Gender Congruent	Per il mio compleanno non avevo organizzato una festa, ma i miei amici mi hanno fatto una$_F$ gradita$_F$ sorpresa$_F$. *For my birthday I had not organized a party, but my friends made me a$_F$ welcomed$_F$ surprise$_F$.*	Per il mio compleanno non avevo organizzato una festa, ma i miei amici mi hanno fatto una$_F$ bella$_F$ sorpresa$_F$. *For my birthday I had not organized a party, but my friends made me a$_F$ nice$_F$ surprise$_F$.*
Gender Incongruent	Non mi piacciono i film che finiscono male, preferisco quelli con una$_F$ gradita$_F$ sorpresa$_F$. *I don't like films that end badly, I prefer those with a$_F$ welcomed$_F$ surprise$_F$.*	Non mi piacciono i film che finiscono male, preferisco quelli con una$_F$ bella$_F$ sorpresa$_F$. *I don't like films that end badly, I prefer those with a$_F$ nice$_F$ surprise$_F$.*

Table 2: Estimates, standard errors, *t* values, and *p* values of the final linear mixed effects model for 250-500 msec. Model: `Voltage ~ Congruence * Predictiveness * Anteriority * Hemisphere + (1 + Congruence * Predictiveness | Subject) + (1 + Congruence * Predictiveness | Item)`

Fixed effect	**Estimate**	**Std. Err.**	***t* value**	**Pr(> |*t*|)**
Intercept	-0.036	0.055	-0.659	.510
Congruence	0.019	0.055	0.347	.729
Predictiveness	0.016	0.055	0.287	.774
Anteriority	-0.245	0.055	-4.442	<.001 ***
Hemisphere	0.180	0.055	3.265	.001 **
Congruence*Predictiveness	0.016	0.055	0.294	.769
Congruence*Anteriority	-0.397	0.055	-7.198	<.001 ***
Predictiveness*Anteriority	-0.130	0.055	-2.350	.019 *
Congruence*Hemisphere	-0.019	0.055	-0.349	.727
Predictiveness*Hemisphere	-0.010	0.055	-0.173	.863
Anteriority *Hemisphere	0.151	0.055	2.732	.006 **
Congruence*Predictiveness*Anteriority	-0.173	0.055	-3.147	.002 **
Congruence*Predictiveness*Hemisphere	0.056	0.055	1.024	.306
Congruence*Anteriority*Hemisphere	0.008	0.055	0.152	.879

Predictiveness*Anteriority*Hemisphere	-0.007	0.055	-0.124	.901
Congruence*Predictiveness*Anteriority*Hemisphere	-0.037	0.055	-0.663	.507

Table 3: Estimates, standard errors, *t* values, and *p* values of the final linear mixed effects model for 500-1000 msec. Model: `Voltage ~ Congruence * Predictiveness * Anteriority * Hemisphere + (1 + Congruence * Predictiveness | Subject) + (1 + Congruence * Predictiveness | Item)`

Fixed effect	**Estimate**	**Std. Err.**	***t* value**	**Pr(> |*t*|)**
Intercept	-0.015	0.062	-0.234	.815
Congruence	0.021	0.062	0.334	.738
Predictiveness	0.002	0.062	0.032	.975
Anteriority	0.005	0.062	0.073	.942
Hemisphere	0.133	0.062	2.141	.032 *
Congruence*Predictiveness	0.011	0.062	0.183	.855
Congruence*Anteriority	-0.372	0.062	-5.974	<.001 ***
Predictiveness*Anteriority	0.088	0.062	1.419	0.156
Congruence*Hemisphere	0.014	0.062	0.221	.825
Predictiveness*Hemisphere	-0.050	0.062	-0.803	.422
Anteriority*Hemisphere	0.036	0.062	0.615	.539
Congruence*Predictiveness*Anteriority	-0.157	0.062	-2.516	.012 *
Congruence*Predictiveness*Hemisphere	-0.018	0.062	-0.293	.769
Congruence*Anteriority*Hemisphere	-0.036	0.062	-0.586	.558

Predictiveness*Anteriority*Hemisphere	0.017	0.062	0.280	.780
Congruence*Predictiveness*Anteriority*Hemisphere	-0.066	0.062	-1.065	.287

Table 4: Estimates, standard errors, *t* values, and *p* values of the final linear mixed effects model for 650-800 msec. Model: `Voltage ~ Congruence * Predictiveness * Anteriority * Hemisphere + (1 + Congruence * Predictiveness | Subject) + (1 + Congruence * Predictiveness | Item)`

Fixed effect	**Estimate**	**Std. Err.**	***t* value**	**Pr(> |*t*|)**
Intercept	-0.007	0.070	-0.096	.923
Congruence	0.028	0.070	0.398	.690
Predictiveness	-0.009	0.070	-0.125	.901
Anteriority	0.198	0.070	2.809	.005 **
Hemisphere	0.230	0.070	3.262	.001 **
Congruence*Predictiveness	0.032	0.070	0.451	.652
Congruence*Anteriority	-0.327	0.070	-4.650	<.001 ***
Predictiveness*Anteriority	0.117	0.070	1.658	.098 .
Congruence*Hemisphere	0.045	0.070	0.634	.526
Predictiveness*Hemisphere	-0.035	0.070	-0.495	.620
Anteriority*Hemisphere	0.045	0.070	0.643	.520
Congruence*Predictiveness*Anteriority	-0.230	0.070	-3.271	.001 **
Congruence*Predictiveness*Hemisphere	-0.017	0.070	-0.236	.813
Congruence*Anteriority*Hemisphere	-0.031	0.070	-0.444	.657

Predictiveness*Anteriority*Hemisphere	-0.003	0.070	-0.049	.961
Congruence*Predictiveness*Anteriority*Hemisphere	-0.066	0.070	-0.939	.348

Table 5: Estimates, standard errors, *t* values, and *p* values of the final linear mixed effects model for ERPs elicited by the article 300-500 msec post-article onset. Model: `Voltage ~ Congruence * Predictiveness * Anteriority * Hemisphere + (1 + Congruence * Predictiveness | Subject) + (1 + Congruence * Predictiveness | Item)`

Fixed effect	**Estimate**	**Std. Err.**	***t* value**	**Pr(> |*t*|)**
Intercept	0.017	0.056	0.311	.756
Congruence	0.024	0.056	0.431	.666
Predictiveness	0.004	0.056	0.074	.941
Anteriority	-0.038	0.056	-0.673	.501
Hemisphere	0.312	0.056	5.565	<.001 ***
Congruence*Predictiveness	0.025	0.056	0.448	.654
Congruence*Anteriority	-0.171	0.056	-3.048	.002 **
Predictiveness*Anteriority	0.102	0.056	1.820	.069 .
Congruence*Hemisphere	0.009	0.056	0.158	.875
Predictiveness*Hemisphere	-0.005	0.056	-0.085	.932
Anteriority*Hemisphere	0.139	0.056	2.470	0.014 *
Congruence*Predictiveness*Anteriority	0.042	0.056	0.751	.452
Congruence*Predictiveness*Hemisphere	-0.046	0.056	-0.821	.411

Congruence*Anteriority*Hemisphere	0.071	0.056	1.267	.205
Predictiveness*Anteriority*Hemisphere	-0.014	0.056	-0.255	.799
Congruence*Predictiveness*Anteriority*Hemisphere	0.002	0.056	0.031	.975

Figure 1: Grand averaged waveforms to target nouns following a locally neutral or predictive adjective under globally congruent or incongruent gender conditions to the target noun at site CP4, low pass filtered at 15 Hz. Voltage maps compare ERPs evoked by the target noun between 250 and 500 msec (predictive – neutral) for congruent and incongruent conditions.

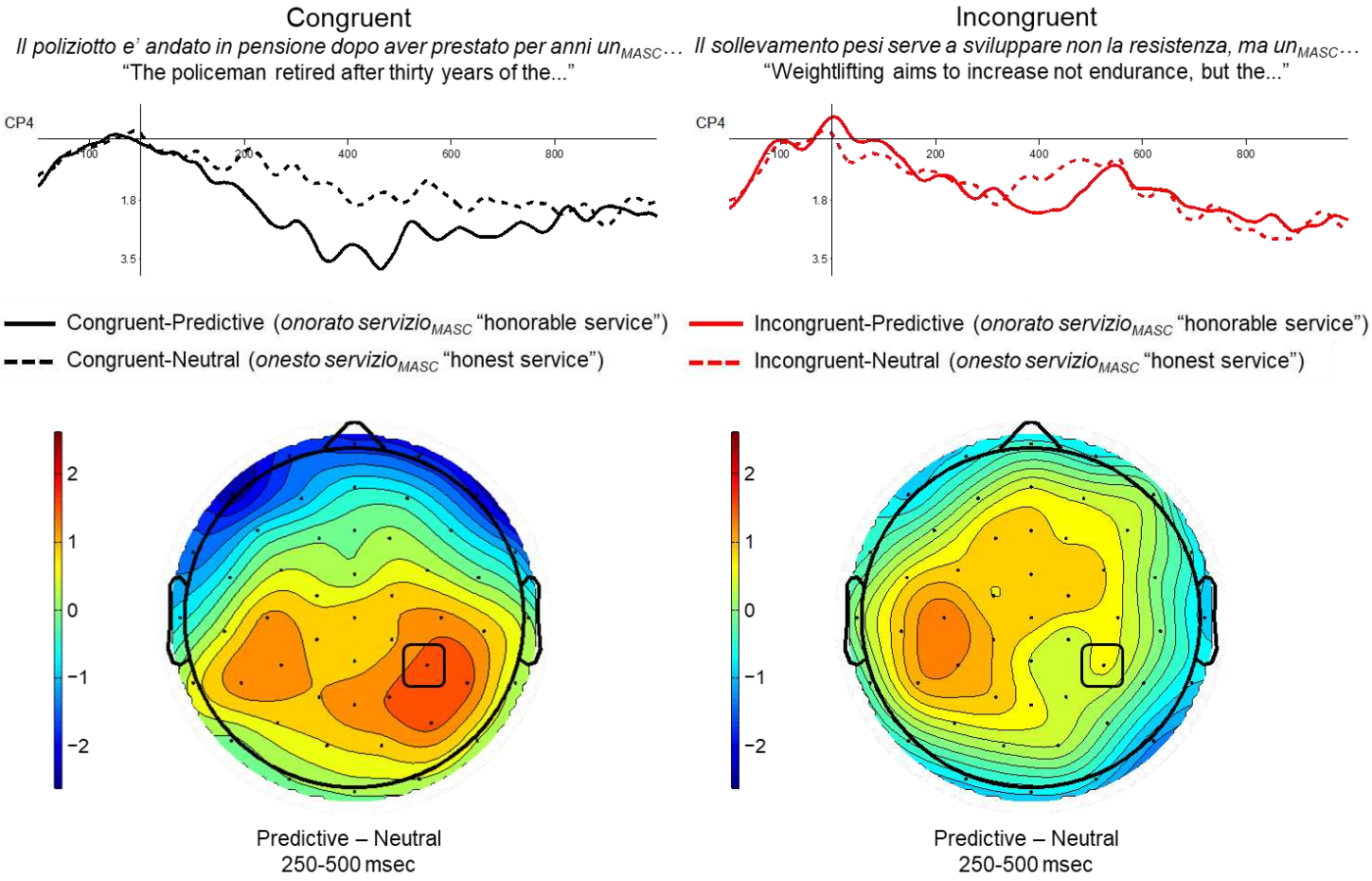


Figure 2: Grand-averaged waveforms to target nouns following a locally predictive or neutral adjective under globally congruent or incongruent gender conditions at site CP4, low-pass filtered at 15 Hz.

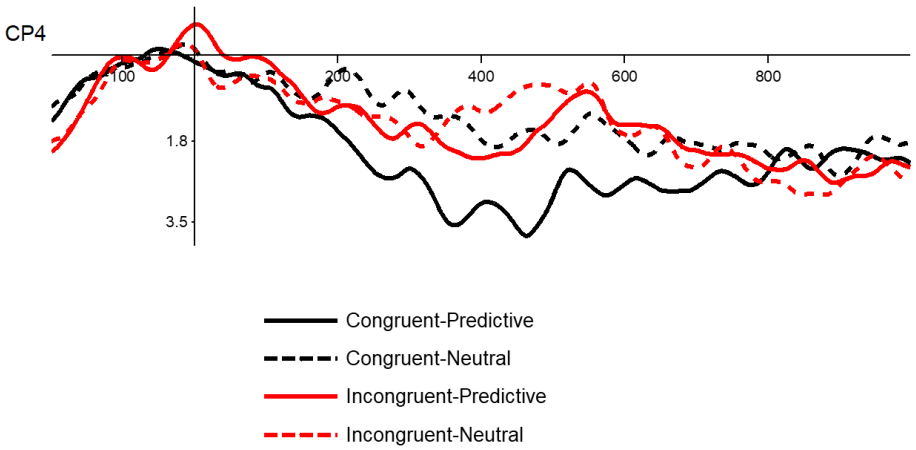


Figure 3: Quadrant analysis of the N400 amplitude during the 250-500 msec time window. Bar plots comparing grand-averaged amplitudes in each of the four quadrants shown on the voltage maps, for Predictiveness under Congruence. Error bars show 95% confidence intervals. Voltage maps comparing average ERP amplitude difference between locally neutral and predictive target nouns between 250-500 msec for globally congruent and incongruent conditions.

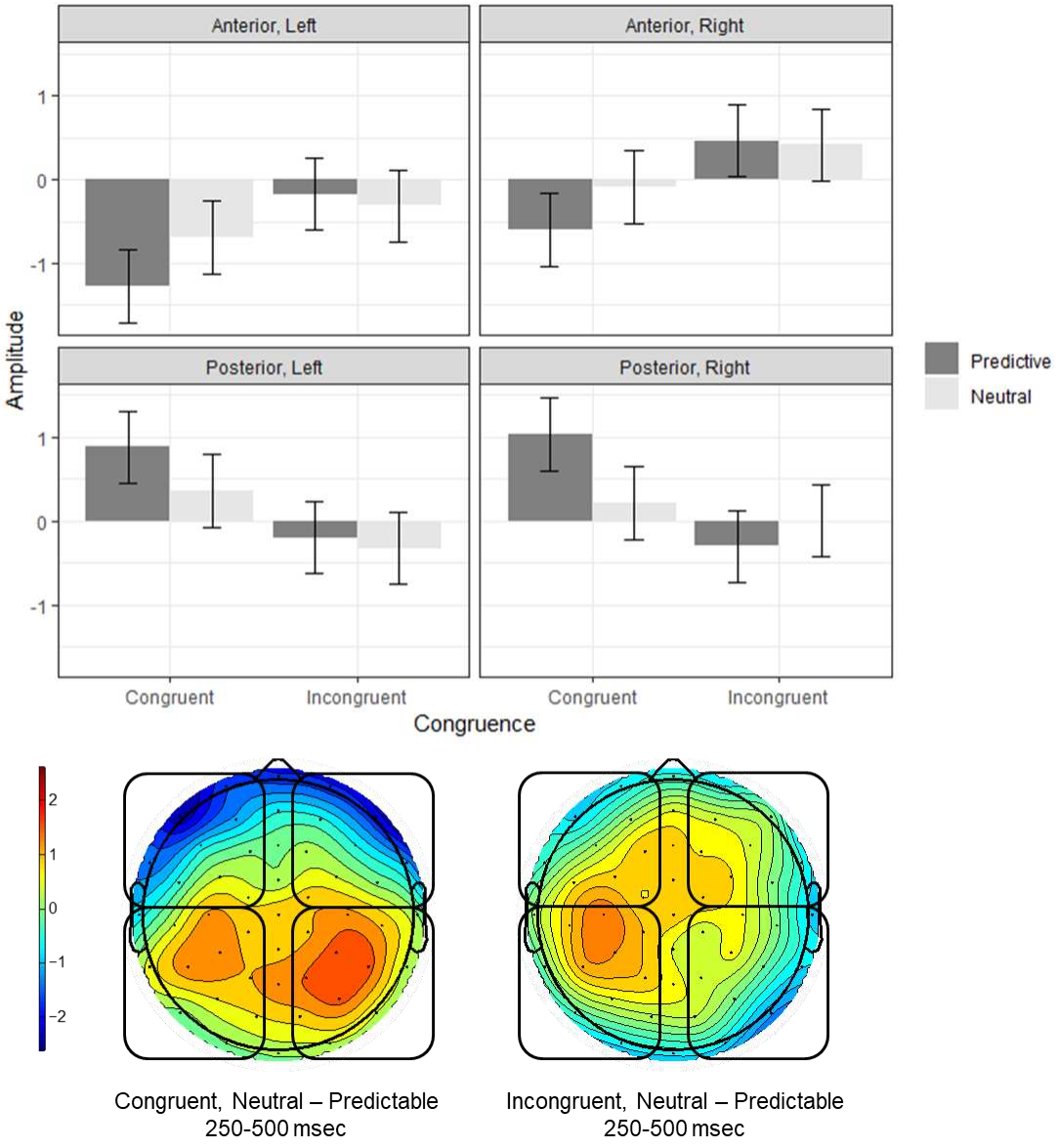


Figure 4: Grand-averaged waveforms to target nouns following a locally predictive or neutral adjective under globally congruent or incongruent gender conditions at site Fz, low-pass filtered at 15 Hz. A) The voltage map compares ERPs evoked by the target noun between 500 and 1000 msec for incongruent minus congruent conditions B) Voltage maps compare ERPs evoked by the target noun between 650 and 800 msec (neutral-predictive) for congruent and incongruent conditions.

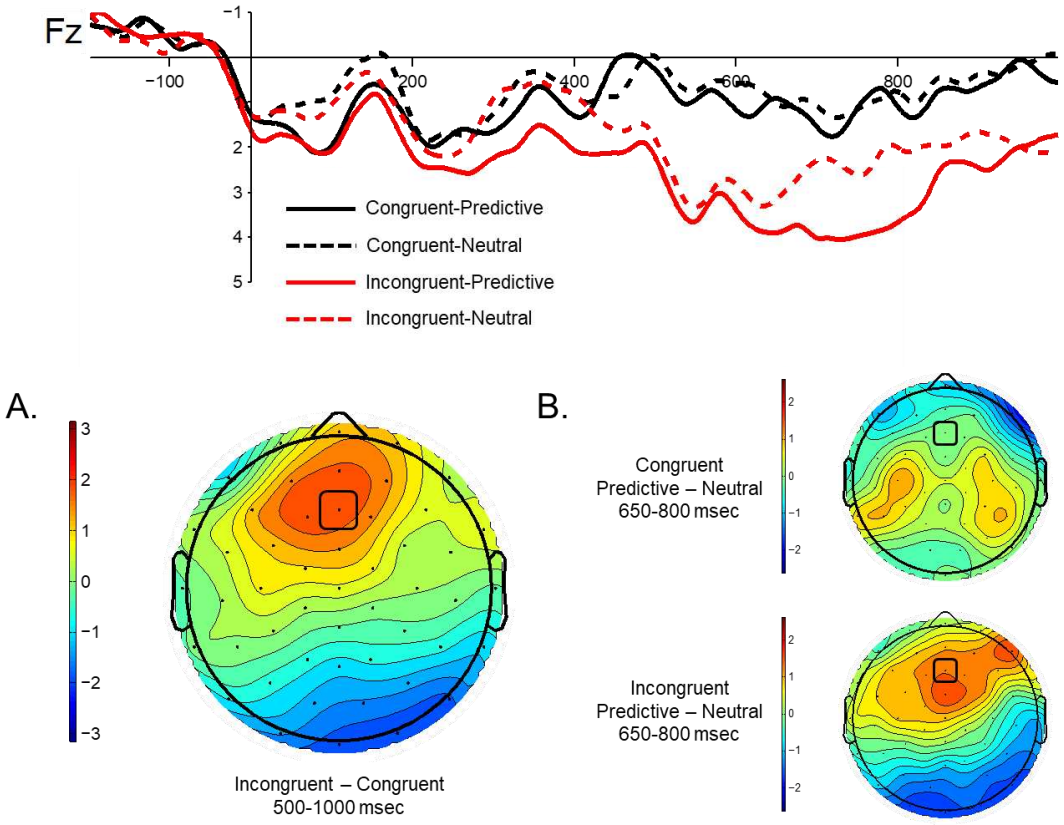


Figure 5. Quadrant analysis of the post-N400 amplitude during the 500-1000 msec time window. Bar plots comparing grand-averaged amplitudes in each of the four quadrants shown on the voltage maps, for Predictiveness under Congruence. Error bars show 95% confidence intervals. Voltage maps comparing average ERP amplitude difference between locally neutral and predictive target nouns between 500-1000 msec for globally congruent and incongruent conditions.

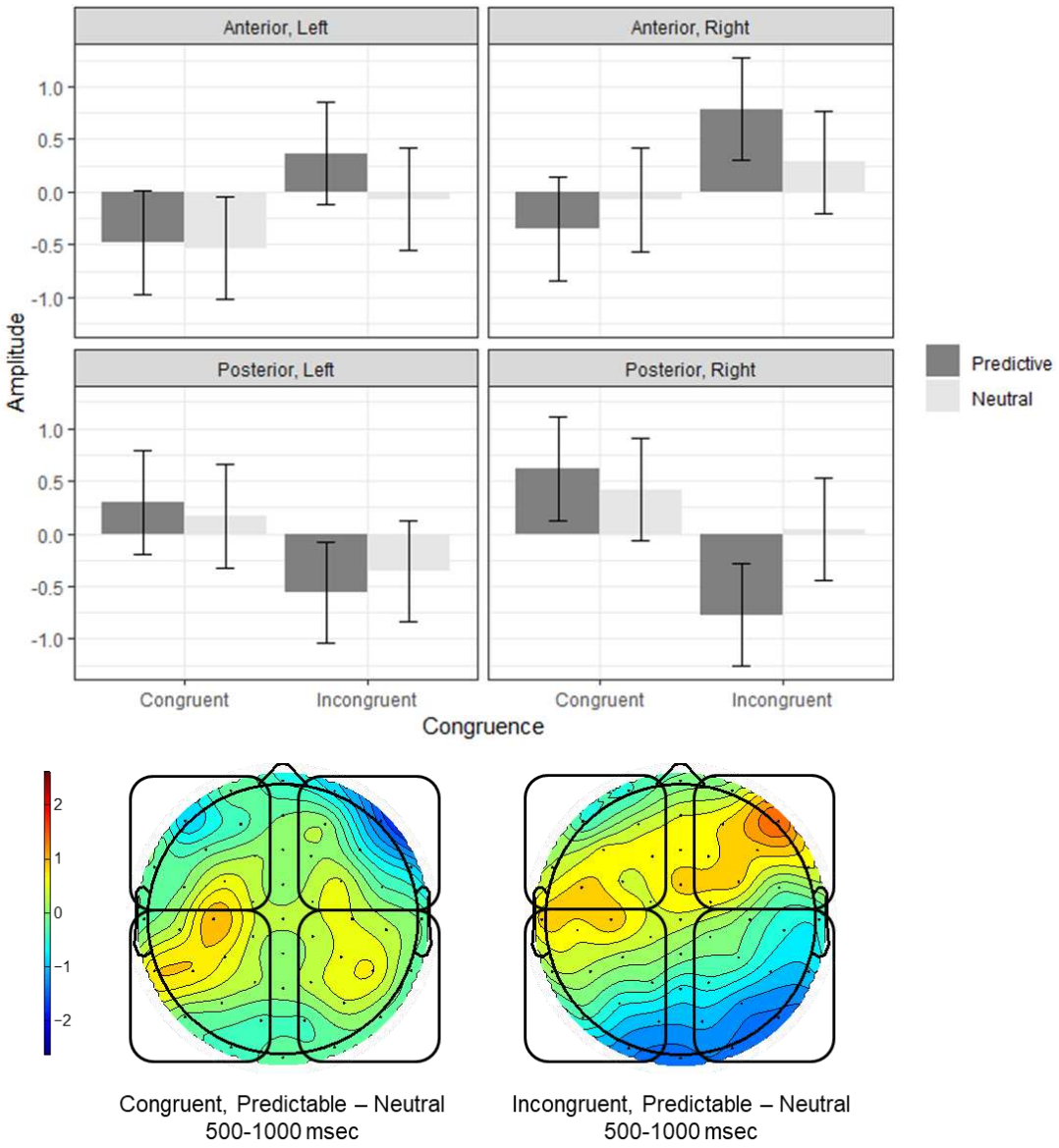


Figure 6. Quadrant analysis of the post-N400 amplitude during the 650-800 msec time window. Bar plots comparing grand-averaged amplitudes in each of the four quadrants shown on the voltage maps, for Predictiveness under Congruence. Error bars show 95% confidence intervals. Voltage maps comparing average ERP amplitude difference between locally neutral and predictive target nouns between 650-800 msec for globally congruent and incongruent conditions.

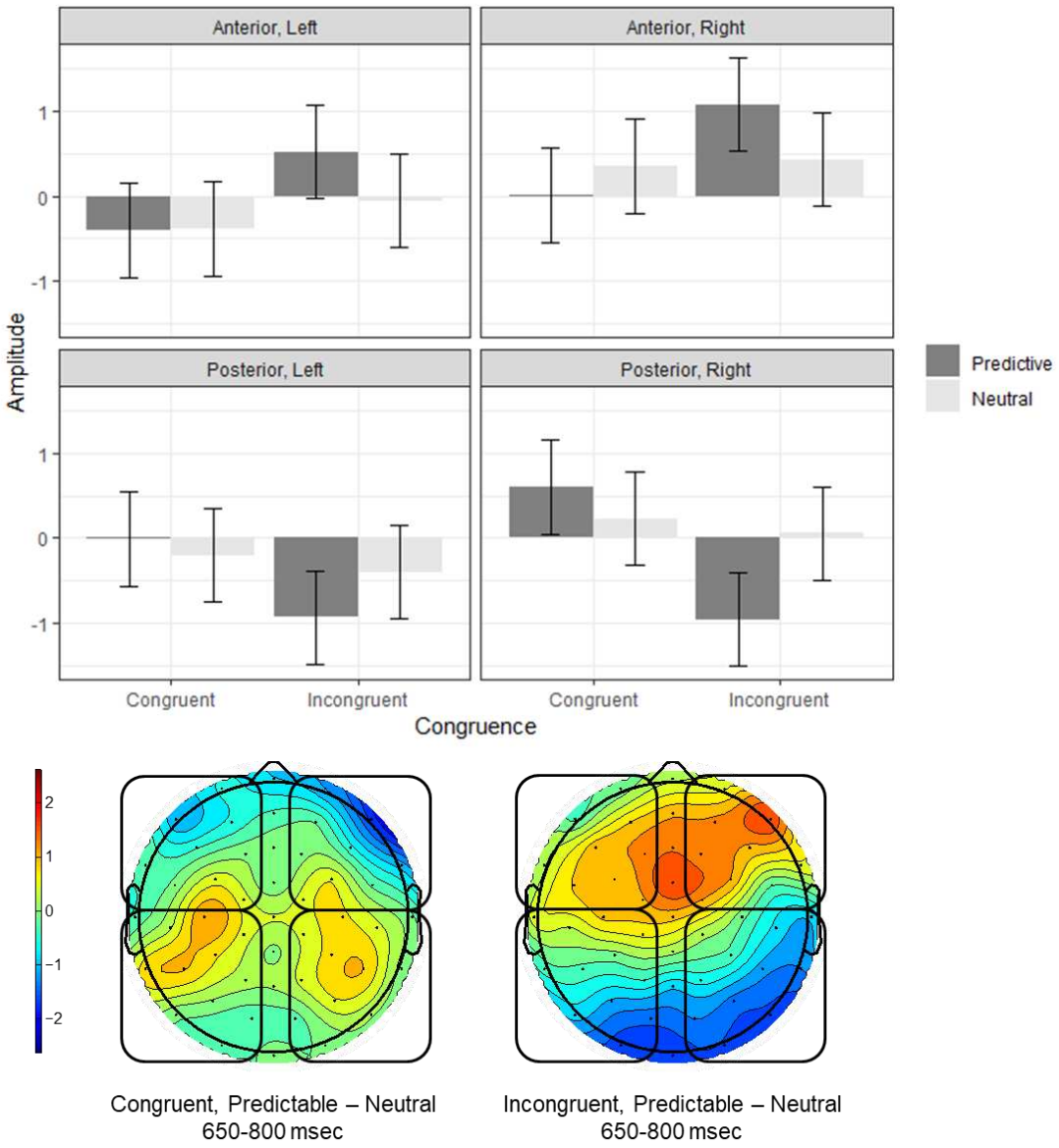
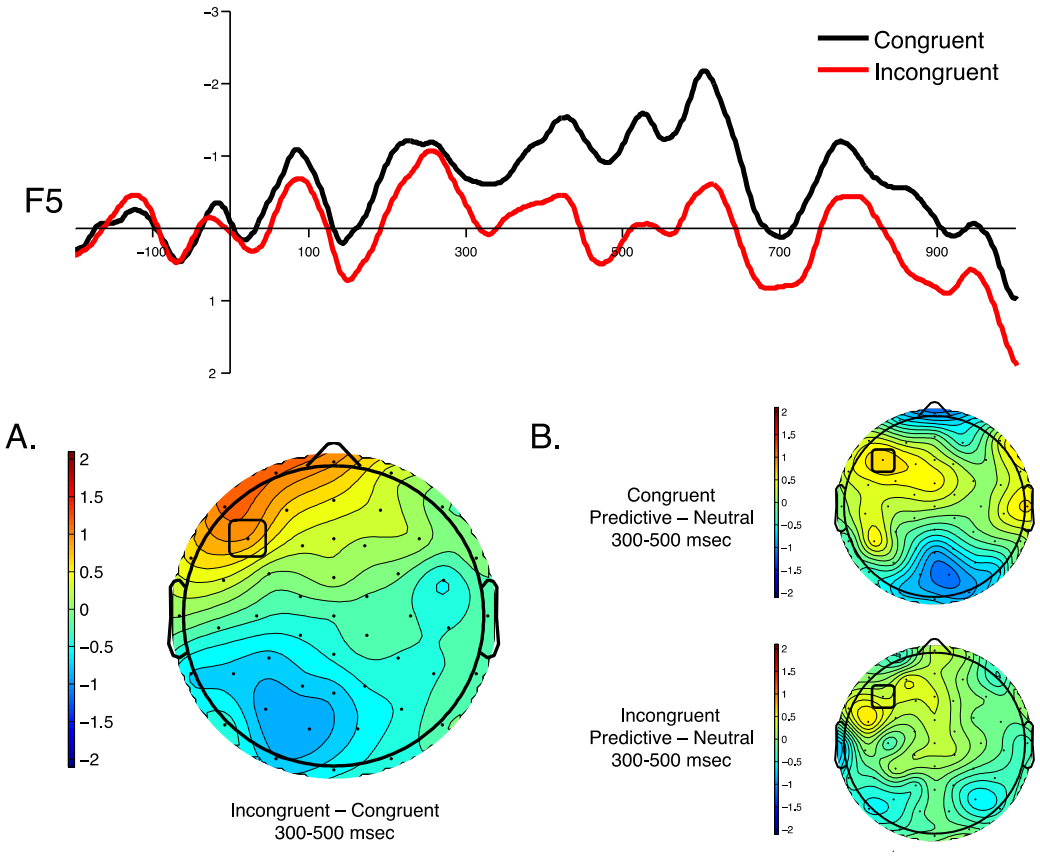


Figure 7. Grand-averaged waveforms to globally gender congruent or incongruent articles at site F5, low-pass filtered at 15 Hz. A) The voltage map compares ERPs evoked by the determiner between 300 and 500 msec for incongruent minus congruent conditions B) Voltage maps compare ERPs evoked by the target noun between 300 and 500 msec (predictive-neutral) for congruent and incongruent conditions.



Appendix A: List of stimuli. The predictiveness of the noun given the adjective is given for both predictive and neutral adjectives under *p(N|A)*. The predictiveness of any noun following the neutral adjective is given under *max p*. The cloze probability for the expected noun given the congruent context is given under *cloze*.

	Congruent Context	**Incongruent Context**	**Det**	**Adjectives**					**Noun**		
				Predictive		**Neutral**			**Expected noun**	**Cloze**	**Gen**
				Adj.	**p(N|A)**	**Adj.**	**p(N|A)**	**max p**			
1	Molti sostengono che il segretario abbia rubato dei soldi, ma lui ha respinto	Prima dell'esecuzione, il condannato ha consumato	la	infamante	0.48	brutta	0	0.11	accusa	98%	F
2	Prima dell'esecuzione, il condannato ha consumato	Molti sostengono che il segretario abbia rubato dei soldi, ma lui ha respinto	un	frugale	0.43	triste	0.0004	0.06	pasto	70%	M

3	Lucia è scivolata giù dalle scale, ma è riuscita a non farsi male durante	Dopo il rapimento, la famiglia dell'ostaggio ha lanciato	la	rovinosa	0.43	breve	0	0.12	caduta	95%	F
4	Dopo il rapimento, la famiglia dell'ostaggio ha lanciato	Lucia è scivolata giù dalle scale, ma è riuscita a non farsi male durante	un	accorato	0.64	insolito	0.01	0.04	appello	71%	M
5	Mi pare che questa sia la strada giusta, ma non so dirlo con	Credevo di non aver dimenticato niente, ma ora mi sta venendo	una	matematica	0.45	completa	0.0006	0.05	certezza	82%	F
6	Credevo di non aver dimenticato niente, ma ora mi sta venendo	Mi pare che questa sia la strada giusta, ma non so dirlo con	un	amletico	0.44	inspiegabile	0	0.04	dubbio	72%	M

7	Il fornitore ha dovuto abbassare i prezzi al minimo per battere	La sposa era già in chiesa, ma doveva ancora arrivare	la	sleale	0.8	attuale	0	0.09	concorrenza	84%	F
8	La sposa era già in chiesa, ma doveva ancora arrivare	Il fornitore ha dovuto abbassare i prezzi al minimo per battere	lo	promesso	0.48	giovane	0.0006	0.07	sposo	75%	M
9	Da quando gli hanno pignorato la casa, Gianni è senza	Il nostro archivio va riordinato, e purtroppo il capo ha assegnato a me	una	fissa	0.79	vera	0.0002	0.03	dimora	31%	F

10	Il nostro archivio va riordinato, e purtroppo il capo ha assegnato a me	Da quando gli hanno pignorato la casa, Gianni è senza	il	ingrato	0.65	importante	0.002	0.03	compito	78%	M
11	Il sollevamento pesi serve a sviluppare non la resistenza, ma	Il poliziotto è andato in pensione, dopo aver prestato per anni	la	erculea	0.5	maggiore	0.009	0.04	forza	53%	F
12	Il poliziotto è andato in pensione, dopo aver prestato per anni	Il sollevamento pesi serve a sviluppare non la resistenza, ma	un	onorato	0.65	onesto	0.007	0.1	servizio	78%	M

13	Da quando è fuggita di casa, Cecilia ha perso ogni contatto con	Laura è sposata da poco, ma litiga già in continuazione con	la	benestante	0.41	vecchia	0.002	0.05	famiglia	54%	F
14	Laura è sposata da poco, ma litiga già in continuazione con	Da quando è fuggita di casa, Cecilia ha perso ogni contatto con	il	fedifrago	0.45	proprio	0	0.14	marito	82%	M
15	All'arrivo della polizia, il rapinatore ha tentato	Il malato prima non riusciva a camminare, oggi per la prima volta ha fatto	una	rocambolesca	0.35	veloce	0.003	0.03	fuga	63%	F

16	Il malato prima non riusciva a camminare, oggi per la prima volta ha fatto	All'arrivo della polizia, il rapinatore ha tentato	un	felpato	0.47	piccolo	0.01	0.05	passo	83%	M
17	Il bradipo viene spesso deriso per	A fine spettacolo, gli attori sono usciti alla ribalta a ricevere	la	esasperante	0.46	incredibile	0.002	0.04	lentezza	75%	F
18	A fine spettacolo, gli attori sono usciti alla ribalta a ricevere	Il bradipo viene spesso deriso per	un	scrosciante	0.84	lungo	0.01	0.13	applauso	85%	M
19	Pur essendo quasi scarica, la torcia emette ancora	Subito dopo l'incidente, diversi passanti si sono fermati a prestare	una	fioca	0.59	forte	0.0002	0.03	luce	96%	F

20	Subito dopo l'incidente, diversi passanti si sono fermati a prestare	Pur essendo quasi scarica, la torcia emette ancora	il	pronto	0.72	necessario	0.0008	0.04	soccorso	85%	M
21	Mario è morto da sei anni ma i familiari ne compiangono ancora	Il diamante è stato tagliato da un rinomato gioelliere, il che ne ha aumentato	la	prematura	0.47	assurda	0.001	0.05	scomparsa	51%	F
22	Il diamante è stato tagliato da un rinomato gioelliere, il che ne ha aumentato	Mario è morto da sei anni ma i familiari ne compiangono ancora	il	inestimabile	0.55	originale	0	0.02	valore	86%	M

23	Le infiltrazioni sono diventate un problema, è ora che l'amministratore risolva	Bisogna educare I figli non solo a parole, ma anche dando	la	annosa	0.41	fastidiosa	0.009	0.05	questione	38%	F
24	Bisogna educare I figli non solo a parole, ma anche dando	Le infiltrazioni sono diventate un problema, è ora che l'amministratore risolva	un	fulgido	0.57	iniziale	0	0.02	esempio	87%	M
25	Tra me e Lucia c'è un problema, ma non ho il coraggio di affrontare	Il palazzo è andato a fuoco, ma non si sa ancora chi abbia appiccato	la	spinosa	0.72	banale	0.01	0.09	questione	50%	F

26	Il palazzo è andato a fuoco, ma non si sa ancora chi abbia appiccato	Tra me e Lucia c'è un problema, ma non ho il coraggio di affrontare	il	doloso	0.69	tremendo	0.008	0.05	incendio	90%	M
27	Ho ripetuto molte volte la domanda a Marco, finchè mi ha dato	Andare in bici non è difficile, basta saper mantenere	una	evasiva	0.44	nuova	0.0003	0.07	risposta	94%	F
28	Andare in bici non è difficile, basta saper mantenere	Ho ripetuto molte volte la domanda a Marco, finchè mi ha dato	un	precario	0.53	buon	0.001	0.09	equilibrio	91%	M

29	Marina ha più di cinquant'anni, alla sua età dovrebbe vestirsi come si conviene a	Mantenere la proprietà pubblica in buono stato è il dovere di	una	distinta	0.42	qualsiasi	0.0002	0.05	signora	53%	F
30	Mantenere la proprietà pubblica in buono stato è il dovere di	Marina ha più di cinquant'anni, alla sua età dovrebbe vestirsi come si conviene a	un	privato	0.6	normale	0.01	0.04	cittadino	92%	M
31	Ieri ho parcheggiato in doppia fila e vigili mi hanno fatto	Avere accesso alle cure mediche non è un lusso, è	una	salatissima	0.59	pesante	0.003	0.04	multa	96%	F

32	Avere accesso alle cure mediche non è un lusso, è	Ieri ho parcheggiato in doppia fila e vigili mi hanno fatto	un	inalienabile	0.52	effettivo	0.003	0.05	diritto	92%	M
33	Il nonno è sempre stato sedentario, e alla lunga questo gli ha danneggiato	Ho comprato delle scarpe nuove, anche se in realtà non ne avevo	la	malferma	0.41	debole	0.0009	0.13	salute	39%	F
34	Ho comprato delle scarpe nuove, anche se in realtà non ne avevo	Il nonno è sempre stato sedentario, e alla lunga questo gli ha danneggiato	un	impellente	0.67	grande	0.0008	0.02	bisogno	94%	M

35	Per il mio compleanno non avevo organizzato una festa, ma i miei amici mi hanno fatto	Non mi piacciono i film che finiscono male, preferisco quelli con	una	gradita	0.53	bella	0.009	0.04	sorpresa	88%	F
36	Non mi piacciono i film che finiscono male, preferisco quelli con	Per il mio compleanno non avevo organizzato una festa, ma i miei amici mi hanno fatto	un	lieto	0.88	diverso	0.0005	0.08	fine	95%	M
37	Ci abbiamo messo tanto ad arrivare perché ci siamo fermati lungo	Trattando con il venditore della casa, siamo riusciti a far abbassare	la	accidentata	0.41	solita	0.002	0.06	strada	59%	F

38	Trattando con il venditore della casa, siamo riusciti a far abbassare	Ci abbiamo messo tanto ad arrivare perché ci siamo fermati lungo	il	modico	0.59	eccessivo	0.002	0.05	prezzo	95%	M
39	La cantante jazz ha cancellato il concerto perché ha perso	Subito dopo il parto, il medico ha tagliato	la	gutturale	0.67	celebre	0.0009	0.03	voce	56%	F
40	Subito dopo il parto, il medico ha tagliato	La cantante jazz ha cancellato il concerto perché ha perso	il	ombelicale	0.96	sanguinolento	0	0.09	cordone	96%	M